

Định danh chủng vi khuẩn tiềm năng chứa gen oxy hoá arsenite *aioA* trong môi trường đất tại Việt Nam

Hoàng Xuân Tùng

Lớp Tin Sinh học 752510, Trường Hoá và Khoa học Sư sống, Đại học Bách khoa Hà Nội

Tóm tắt: Việc xác định các chủng vi sinh vật có khả năng oxy hóa arsenite đóng vai trò quan trọng trong nghiên cứu xử lý ô nhiễm kim loại nặng tại Việt Nam. Trong nghiên cứu này, một chủng vi khuẩn tiềm năng đã được phân lập từ môi trường đất và định danh dựa trên trình tự gen 16S rRNA. Phân tích BLASTN và phát sinh loài cho thấy chủng này có quan hệ gần gũi với *Ancylobacter novellus* và thuộc chi *Ancylobacter*. Để kiểm tra sự hiện diện của gen *aioA*, một gen chức năng liên quan đến khả năng oxy hóa arsenite, các trình tự ortholog từ các loài gần gũi đã được thu thập và phân tích nhằm xác định vùng bảo tồn. Một cặp mồi đặc hiệu đã được thiết kế và đánh giá về tính đặc hiệu cũng như các đặc tính lý hoá. Ngoài ra, protein được mã hoá bởi gen *aioA* cũng được phân tích đặc điểm sinh hoá và dự đoán cấu trúc không gian ba chiều. Những kết quả thu được nhằm tạo nền tảng cho việc sàng lọc hiệu quả các chủng vi sinh vật tiềm năng và mở ra triển vọng ứng dụng chúng trong các chiến lược xử lý arsenite ngoài thực địa một cách chính xác và bền vững.

Từ khoá: 16S rRNA, *aioA*, phát sinh loài, thiết kế mồi.

I. MỞ ĐẦU

Các loài *Ancylobacter* sp. trước đây được phân loại trong chi *Microcyclus* của họ *Spirosomaceae* [1], chúng là những vi khuẩn hình que cong, hiếu khí bắt buộc, Gram âm. Tuy nhiên, những tiến bộ gần đây trong công nghệ định danh sinh vật và phân tích phát sinh loài đã mang lại những hiểu biết mới về đặc điểm di truyền và phân loại của chi này. Phân tích phát sinh loài dựa trên trình tự 16S rRNA đã chỉ ra rằng các các loài thuộc chi *Ancylobacter* là thành viên của lớp *Alphaproteobacteria* [2]. Các thành viên thuộc chi *Ancylobacter* thường được tìm thấy trong môi trường đất và nước ngọt. Trên môi trường rắn, khuẩn lạc của chúng có màu trắng kem; ở môi trường lỏng, chúng hình thành màng nổi. Các chủng đều dương tính với phản ứng oxidase và catalase [3]. *Ancylobacter* là nhóm vi khuẩn dị dưỡng hóa học, methylotrophic tùy nghi, có khả năng sử dụng đa dạng các loại đường và muối của acid hữu cơ làm nguồn carbon, một số chủng còn thể hiện khả năng hóa đường vô cơ khi sử dụng hydro phân tử [4]. Hiện tại, chi *Ancylobacter* bao gồm hơn 22 loài được công bố hợp lệ (<https://lpsn.dsmz.de/genus/ancylobacter>). Nghiên cứu gần đây đã nhấn mạnh ứng dụng tiềm năng của các loài *Ancylobacter* như khả năng khử halogen [5], phân hủy các hợp chất hữu cơ như 1,2-dichloroethane [6] và 2-chloroethylvinylether [7], sản xuất cellulose năng suất cao [8] và oxi hoá arsenite [9], [10].

Ở điều kiện tự nhiên, arsenite (AsIII) là một trong hai dạng tồn tại chính của arsenic trong môi trường. Loại As(III) trên đã được chứng minh là gây độc hại đến các quá trình trao đổi chất của sinh vật sống như ức chế hệ thống enzyme, ức chế sự sửa chữa DNA và hô hấp ty thể, gây ra nhiều phản ứng cảng thẳng, thúc đẩy khuếch đại gen và tạo tổn thương di truyền [11]. Trong khi đó, đối với nhiều chủng vi khuẩn oxy hóa arsenite dị dưỡng và hoá dưỡng, quá trình oxy hóa As(III) thông qua hệ thống enzyme Aio (gồm hai tiểu đơn vị AioA và AioB với gen mã hoá các tiểu đơn vị tương ứng là *aioA*, *aioB*) [12], [13] lại được coi là một cơ chế giải độc, một nguồn năng lượng bổ sung để phát triển khi có carbon dioxide trong điều kiện hiếu khí [14]–[17] và khử nitrat [18]. Đồng thời với việc gen *aioA* đã được sử dụng để làm chất đánh dấu phân tử để khảo sát sự đa dạng và phân loại các cộng đồng vi sinh vật oxy hóa As(III) trong môi trường [19]–[21], vi sinh vật đã được tạo điều kiện mạnh mẽ để trở thành những ứng cử viên tiềm năng cho việc phân huỷ arsenite trong các khu vực ô nhiễm với hiệu quả cao, quy mô lớn với chi phí tiết kiệm. Cho đến nay, các gen tương đồng mã hóa AioAB đã được tìm thấy đa dạng ở các chủng thuộc Actinobacteria, Aquificae, Bacteroidetes, Chlorobi, Chloroflexi, Crenarchaeota, Deinococcus-Thermus, Firmicutes, Nitrospira và α -, β -, γ -Proteobacteria [13]. Mục tiêu của bài báo cáo này là định danh chủng vi khuẩn được cho là thuộc chi *Ancylobacter* dựa trên gen mã hoá 16S rRNA và thiết kế mồi cho phản ứng PCR để xác định sự có mặt của gen *aioA* trong chủng vi sinh vật này.

II. NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

A. Định danh chủng vi sinh vật

Chủng vi sinh vật đã được phân lập từ môi trường tự nhiên với trình tự gen được xác định thông qua phương pháp giải trình tự Sanger hai chiều. Trình tự consensus của đoạn gen mã hoá 16S rRNA được thiết kế thông qua phần mềm MEGA12 phiên bản 12.0.11 [22] và được khuếch đại bằng cặp mồi 27F/1492R [23]. Phân tích BLASTN với cơ sở dữ liệu trình tự rRNA/ITS/16S rRNA của vi khuẩn và cổ khuẩn đã loại trừ các trình tự thu thập từ môi trường [24] được sử dụng để xác định các loài có quan hệ gần gũi.

B. Xây dựng cây phát sinh loài dựa trên gen 16S rRNA

Vị trí phân loại của chủng được xác định dựa trên cơ sở dữ liệu Taxonomy của NCBI [25]. Để làm sáng tỏ mối quan hệ phát sinh loài giữa chủng C3 và các loài liên quan, một cây phát sinh loài được xây dựng trên trình tự 16S rRNA. Phân tích bao gồm các loài thuộc chi *Ancylobacter*, cũng như các chi

liên quan của nó là *Aquabacter*, *Xanthobacter*, *Azorhizobium* và *Labrys*. Bộ dữ liệu gen 16S rRNA đã được căn chỉnh bằng CLUSTALW [26] và các mối quan hệ phát sinh loài đã được suy ra bằng phương pháp neighbor-joining (NJ) [22]. Cấu trúc tòpô của cây đã được đánh giá bằng phương pháp bootstrap với 1000 lần lây mẫu lặp lại, với *Escherichia coli* VANM 2 được sử dụng làm nhóm ngoài (outgroup).

C. Thiết kế mồi khuếch đại gen aioA

Để kiểm tra sự hiện diện của gen chức năng ở chủng vi sinh vật được phân lập, cặp mồi PCR đã được thiết kế dựa trên vùng bảo thủ của các ortholog có chức năng quan tâm được chú thích từ các loài có quan hệ gần gũi. Các trình tự ortholog thuộc họ Xanthobacteraceae (Bảng I), bao gồm các loài trong chi *Rhizobium*, *Shinella* và *Sinorhizobium*, đã được thu thập từ cơ sở dữ liệu NCBI GenBank. Trong đó, *Ancylobacter* sp. TS-1 được chọn làm chủng tham chiếu gần gũi nhất về mặt phân loại với chủng nghiên cứu. Các trình tự thu được đã được căn chỉnh bằng cách sử dụng chương trình CLUSTAL OMEGA của EBI [27] và phân tích đa trinh tự nhằm xác định các vùng bảo tồn trong gen *aioA* giữa các chủng vi khuẩn. Sau khi xác định được các vùng bảo tồn, cặp mồi đã được thiết kế bằng công cụ NCBI Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast>) [28].

BẢNG I

TÁM TRÌNH TỰ ORTHOLOG VỚI *Ancylobacter* ĐƯỢC SỬ DỤNG ĐỂ SO SÁNH VÀ TỪ ĐÓ XÁC ĐỊNH ĐƯỢC VÙNG BẢO THỦ.

Tên chủng	Số hiệu truy cập
<i>Ancylobacter</i> sp. TS-1	KX250215.1
<i>Rhizobium</i> sp. YE2-4	KT992350.1
<i>Rhizobium</i> sp. K3	KT992346.1
<i>Rhizobium</i> sp. 3AM-13	KT992341.1
<i>Rhizobium</i> sp. Cug7	MF621580.1
<i>Rhizobium</i> sp. Cug6	MF621579.1
<i>Shinella</i> sp. C23	KT992342.1
<i>Shinella</i> sp. Cug3	MF621576.1
<i>Sinorhizobium</i> KGO-5	AB915361.1

Trình tự protein được chuyển đổi từ trình tự gen mã hóa thông qua công cụ Translate của EXPASY [29] và được tính toán khối lượng phân tử cũng như điểm đẳng điện thông qua công cụ Compute pI/Mw EXPASY [30]–[32]. Xác định sự có mặt của đoạn trình tự tín hiệu cho việc vận chuyển qua màng tế bào của protein sau dịch mã thông qua chương trình SIGNALP 6.0 [33]. Khả năng gắn màng của protein được kiểm tra bằng công cụ DEEPMHMM 1.0 [34]. Cấu trúc thực nghiệm của protein được xác minh thông qua đối chiếu với cơ sở dữ liệu Protein Data Bank [35] và được dự đoán bằng mô hình ALPHAFOLD [36] dựa vào chuỗi trình tự thu được từ đoạn gen mã hóa.

III. KẾT QUẢ VÀ THẢO LUẬN

A. Phân tích cây phát sinh loài dựa trên gen 16S rRNA

Kết quả ở Bảng II cho thấy trình tự 16S rRNA của chủng C3 thể hiện mức độ tương đồng cao nhất (99.25%) với *A. novellus*, tiếp theo là *A. koreensis* (99.10%), *A. lacus*

(97.97%), *A. defluvii* (97.75%) và *A. plantiphilus* (97.67%). Dựa vào cơ sở dữ liệu Taxonomy, vị trí phân loại của chủng nghiên cứu được xác định là thuộc bộ *Hyphomicrobiales*, họ *Xanthobacteraceae*, chi *Ancylobacter*. Như thể hiện trong Hình 1, cây phát sinh loài dựa trên trình tự 16S rRNA chỉ ra rằng các chủng thuộc chi *Ancylobacter* tạo thành một nhánh riêng biệt, với chúng thể hiện mối quan hệ họ hàng gần gũi với *Aquabacter* và *Xanthobacter* hơn là với *Escherichia*. Trong nhánh *Ancylobacter*, chủng C3 có quan hệ họ hàng gần với *A. novellus* DSM 506 và *A. koreensis* NBRC 100963.

BẢNG II

SO SÁNH ĐỘ TƯƠNG ĐỒNG TRÌNH TỰ GEN 16S rRNA CỦA CHỦNG C3 VỚI CÁC CHỦNG KHÁC THUỘC CHI *Ancylobacter* SỬ DỤNG BLASTN.

Tên loài/chủng	Tỷ lệ tương đồng (%)	Số hiệu truy cập
<i>A. novellus</i> DSM 506	99.25	NR_074219.1
<i>A. koreensis</i>	99.10	NR_113962.1
<i>A. koreensis</i>	99.10	NR_041013.1
<i>A. novellus</i> DSM 506	98.50	NR_025859.1
<i>A. lacus</i>	97.97	NR_180559.1
<i>A. defluvii</i>	97.75	NR_133695.1
<i>A. plantiphilus</i>	97.67	NR_180560.1
<i>A. rudongensis</i>	97.68	NR_029047.1
<i>A. dichloromethanicus</i>	97.60	NR_116379.1
<i>A. oerskovii</i>	97.52	NR_042655.1

Việc xác định chính xác vị trí phân loại của chủng vi sinh vật là cơ sở quan trọng cho việc suy luận các đặc tính sinh học liên quan, đặc biệt trong bối cảnh nhiều loài thuộc chi này đã được ghi nhận có khả năng oxy hóa arsenite trong môi trường. Dữ liệu phát sinh loài cũng hỗ trợ việc lựa chọn các chủng tham chiếu phù hợp để phân tích gen chức năng.

B. Đánh giá đoạn mồi khuếch đại gen aioA

Phân tích đa trinh tự cho thấy hai vùng bảo tồn nổi bật nằm tại vị trí nucleotide 103–122 và 403–422 trên trình tự gen *aioA* của *Ancylobacter* sp. TS-1, tương ứng với các đoạn trình tự GTGCACATCGTCATCAAGCC và GGCAAGCTCTATTGTTGGGGC. Dựa trên hai vùng này, một cặp mồi đã được thiết kế gồm mồi xuôi GTGCACATCGTCATCAA và mồi ngược GCCCCAAAATAGAGCTTG (Bảng III). Cặp mồi này đã được kiểm tra lại bằng công cụ BLASTN để đảm bảo tính đặc hiệu cao đối với gen *aioA*, với các thông số như nhiệt độ nóng chảy phù hợp, hàm lượng GC nằm trong khoảng tối ưu và không hình thành cấu trúc thứ cấp.

Chiến lược thiết kế mồi dựa trên ortholog giúp mở rộng khả năng áp dụng cặp mồi với các chủng vi khuẩn khác trong họ Xanthobacteraceae, đặc biệt có lợi trong các khảo sát đa dạng vi sinh vật môi trường, nơi trình tự gen đích có thể biến đổi nhẹ giữa các loài. Tuy nhiên, nghiên cứu vẫn cần thực nghiệm PCR và giải trình tự để xác minh sự hiện diện và hoạt động của gen trong chủng C3.

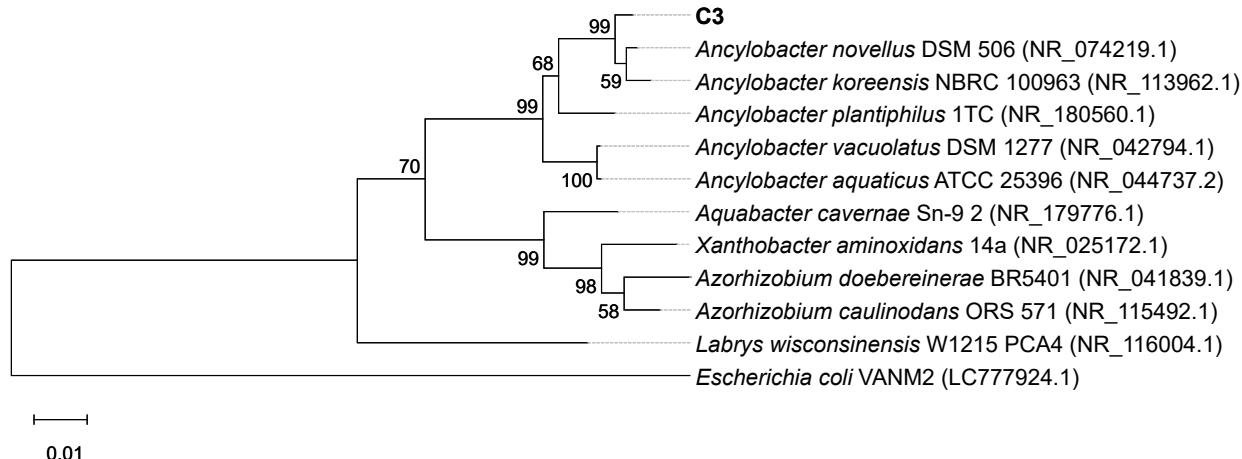
C. Phân tích đặc điểm protein được mã hóa bởi gen aioA

Protein được dịch mã từ gen mã hóa trên có khối lượng phân tử khoảng 35444.57 Mw, điểm đẳng điện pI khoảng 5.08. Các

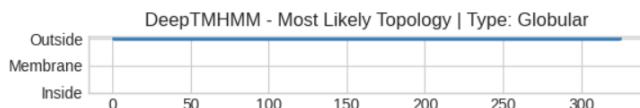
BẢNG III

THÔNG SỐ ĐOẠN MỒI ĐƯỢC THIẾT KẾ DỰA TRÊN CHỦNG THAM CHIỀU *Ancylobacter* sp. TS-1 THÔNG QUA PRIMER-BLAST. CHIỀU DÀI SẢN PHẨM: 360.

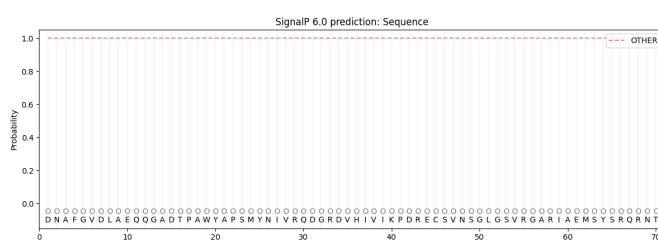
	Sequence (5' - 3')	Template strand	Length	Start	Stop	Tm	GC%	Self Comp.	Self 3' Comp.
Forward primer	GTGCACATCGTCATCAA	Plus	17	103	119	52.32	47.06	6.00	0.00
Reverse primer	GCCCCAAAATAGAGCTTG	Minus	18	422	405	53.34	50.00	4.00	2.00



Hình 1. Cây phát sinh loài theo phương pháp NJ dựa trên trình tự gen 16S rRNA, mô tả mối quan hệ phát sinh loài giữa chủng C3 và các loài gần gũi. Các con số tại các nút nhánh biểu thị giá trị bootstrap 1000 lần lặp lại, chỉ hiển thị các giá trị > 50. Thang tỷ lệ biểu thị 0.01 biến đổi trên mỗi vị trí nucleotide.



Hình 2. Khảo sát khả năng gắn trên màng tế bào của protein bằng DeepTMHMM-1.0.



Hình 3. Khảo sát tín hiệu cho việc vận chuyển qua màng tế bào của protein bằng công cụ SignalP 6.0. Quá trình cải biến sau dịch mã xảy ra ở một số loại protein và chuyển hóa tiền-protein thành dạng có hoạt tính hoặc vận chuyển protein đến vị trí mà tại đó nó có hoạt tính xúc tác (ví dụ như trên màng tế bào, hoặc trong periplasmic). Một trong các quá trình cải biến protein sau dịch mã là vận chuyển protein qua màng tế bào, thông thường, các protein này có một đoạn trình tự peptide làm tín hiệu, sau khi protein được vận chuyển qua màng ra môi trường hoặc ra periplasmic, đoạn trình tự này bị cắt bỏ.

protein này không gắn trên màng tế bào (Hình 2), không có tín hiệu cho việc vận chuyển qua màng tế bào (Hình 3) và có cấu trúc dự đoán như Hình 4.

Dù chưa có xác minh thực nghiệm về hoạt tính enzyme, cấu trúc và đặc điểm sinh hóa protein gợi ý rằng sản phẩm mã hóa có thể đóng vai trò oxy hóa arsenite tương tự các enzyme AioA đã biết. Điều này hỗ trợ giả thuyết rằng chủng C3 có tiềm năng tham gia quá trình chuyển hóa As(III) trong điều kiện tự nhiên hoặc ứng dụng xử lý môi trường.

IV. KẾT LUẬN

Nghiên cứu này đã định danh thành công một chủng vi khuẩn tiềm năng thuộc chi *Ancylobacter* thông qua phân tích trình tự gen 16S rRNA và cây phát sinh loài. Chủng này cho thấy mối quan hệ họ hàng gần gũi với *Ancylobacter novellus*, mở ra khả năng ứng dụng trong các nghiên cứu môi trường liên quan đến arsenic. Trên cơ sở các trình tự ortholog của gen *aioA* thu thập từ các loài vi khuẩn có quan hệ gần gũi, bài nghiên cứu đã xác định được các vùng bảo tồn và thiết kế một cặp mồi PCR đặc hiệu nhằm phát hiện sự hiện diện của gen *aioA* trong chủng vi sinh vật đã phân lập. Đồng thời, phân tích tin sinh học đối với sản phẩm protein được mã hóa đã cung cấp thêm thông tin về đặc điểm sinh hóa và cấu trúc không gian ba chiều tiềm năng của enzyme AioA.

Kết quả nghiên cứu không chỉ cung cấp dữ liệu định danh ban đầu mà còn đặt nền tảng cho việc sàng lọc các chủng vi khuẩn có khả năng oxy hóa arsenite tại Việt Nam. Qua đó, nghiên cứu góp phần định hướng ứng dụng vi sinh vật trong các chiến lược xử lý ô nhiễm arsenite ngoài thực địa theo hướng hiệu quả, chính xác và bền vững trong tương lai.



Hình 4. Cấu trúc ba chiều dự đoán của protein được tạo bởi ALPHAFOLD. Mô hình cho thấy cách gấp không gian, cấu trúc xoắn α , tấm β và tổ chức các miền của protein. Mức độ tin cậy của từng vùng trong cấu trúc được mã hóa màu dựa trên điểm pLDDT, các vùng màu xanh dương thể hiện độ tin cậy cao.

LỜI CẢM ƠN

Em xin gửi lời cảm ơn sâu sắc tới PGS.TS. Nguyễn Lan Hương và TS. Đàm Thúy Hằng vì đã thiết kế các bài thi nghiệm giàu tính ứng dụng, đồng thời tận tình hướng dẫn em trong quá trình tìm hiểu và sử dụng các công cụ tin sinh học hiện đại trong bài báo cáo này.

PHỤ LỤC

Tài liệu bổ sung về các đoạn trình tự gen cũng như các thao tác làm bài được ghi chép trong tệp Word đính kèm.

TÀI LIỆU THAM KHẢO

- [1] J. Ørskov, *Beschreibung eines neuen mikroben, microcyclus aquaticus*, 1928.
- [2] C. Woese, S. Maloy, L. Mandelco, and H. Raj, “Phylogenetic placement of the spirosomaceae,” *Systematic and applied microbiology*, vol. 13, no. 1, pp. 19–23, 1990.
- [3] J. M. LARKIN, P. M. WILLIAMS, and R. TAYLOR, “Taxonomy of the genus *microcyclus* Ørskov 1928: Reintroduction and emendation of the genus *spirosoma* migula 1894 and proposal of a new genus, *flectobacillus*,” *International Journal of Systematic and Evolutionary Microbiology*, vol. 27, no. 2, pp. 147–156, 1977.
- [4] B. Namsaraev and A. Nozhevnikova, “Autotrophic growth of *microcyclus-aquaticus* in an atmosphere of hydrogen,” *Microbiology*, vol. 47, no. 3, pp. 315–318, 1978.
- [5] J. Firsova, N. Doronina, E. Lang, C. Spröer, S. Vuilleumier, and Y. Trotsenko, “*Ancyllobacter dichloromethanicus* sp. nov.—a new aerobic facultatively methylotrophic bacterium utilizing dichloromethane,” *Systematic and Applied Microbiology*, vol. 32, no. 4, pp. 227–232, 2009.
- [6] Z. Wu, Z. Zuo, X. Zhang, et al., “Characterization of a novel 1, 2-dichloroethane degrader *ancylobacter* sp. j3 and use of its immobilized cells in the treatment of polluted groundwater,” *Journal of Hazardous Materials*, p. 138 346, 2025.
- [7] A. J. van den Wijngaard, J. Prins, A. J. Smal, and D. B. Janssen, “Degradation of 2-chloroethylvinylether by *ancylobacter aquaticus* ad25 and ad27,” *Applied and environmental microbiology*, vol. 59, no. 9, pp. 2777–2783, 1993.
- [8] S.-M. Martirani-Von Abercron, D. Pacheco-Sánchez, I. Castillo-Rodríguez, et al., “Spontaneous loss of quorum sensing control selects a new high cellulose producing *ancylobacter* strain,” *International Journal of Biological Macromolecules*, vol. 304, p. 140 620, 2025.
- [9] V. Andreoni, R. Zanchi, L. Cavalca, A. Corsini, C. Romagnoli, and E. Canzi, “Arsenite oxidation in *ancylobacter dichloromethanicus* as3-1b strain: Detection of genes involved in arsenite oxidation and co 2 fixation,” *Current microbiology*, vol. 65, pp. 212–218, 2012.
- [10] F. Battaglia-Brunet, M.-C. Dictor, F. Garrido, et al., “An arsenic (iii)-oxidizing bacterial population: Selection, characterization, and performance in reactors,” *Journal of applied microbiology*, vol. 93, no. 4, pp. 656–667, 2002.
- [11] E. T. Snow, Y. Hu, C. C. Yan, and S. Chouchane, “Modulation of dna repair and glutathione levels in human keratinocytes by micromolar arsenite,” in *Arsenic Exposure and Health Effects III*, Elsevier, 1999, pp. 243–251.
- [12] R. Mukhopadhyay, B. P. Rosen, L. T. Phung, and S. Silver, “Microbial arsenic: From geocycles to genes and enzymes,” *FEMS microbiology reviews*, vol. 26, no. 3, pp. 311–325, 2002.

- [13] S. Yamamura and S. Amachi, "Microbiology of inorganic arsenic: From metabolism to bioremediation," *Journal of bioscience and bioengineering*, vol. 118, no. 1, pp. 1–9, 2014.
- [14] C. G. Bryan, M. Marchal, F. Battaglia-Brunet, *et al.*, "Carbon and arsenic metabolism in thiomonas strains: Differences revealed diverse adaptation processes," *BMC microbiology*, vol. 9, pp. 1–12, 2009.
- [15] K. Duquesne, A. Lieutaud, J. Ratouchniak, D. Muller, M.-C. Lett, and V. Bonnefoy, "Arsenite oxidation by a chemoautotrophic moderately acidophilic thiomonas sp.: From the strain isolation to the gene study," *Environmental microbiology*, vol. 10, no. 1, pp. 228–237, 2008.
- [16] E. Garcia-Dominguez, A. Mumford, E. D. Rhine, A. Paschal, and L. Y. Young, "Novel autotrophic arsenite-oxidizing bacteria isolated from soil and sediments," *FEMS microbiology ecology*, vol. 66, no. 2, pp. 401–410, 2008.
- [17] J. M. Santini, L. I. Sly, R. D. Schnagl, and J. M. Macy, "A new chemolithoautotrophic arsenite-oxidizing bacterium isolated from a gold mine: Phylogenetic, physiological, and preliminary biochemical studies," *Applied and environmental microbiology*, vol. 66, no. 1, pp. 92–97, 2000.
- [18] R. S. Oremland, S. E. Hoeft, J. M. Santini, N. Bano, R. A. Hollibaugh, and J. T. Hollibaugh, "Anaerobic oxidation of arsenite in mono lake water and by a facultative, arsenite-oxidizing chemoautotroph, strain mlhe-1," *Applied and environmental microbiology*, vol. 68, no. 10, pp. 4795–4802, 2002.
- [19] J. Zhang, S. Zhao, Y. Xu, *et al.*, "Nitrate stimulates anaerobic microbial arsenite oxidation in paddy soils," *Environmental science & technology*, vol. 51, no. 8, pp. 4377–4386, 2017.
- [20] S. Xue, X. Jiang, C. Wu, *et al.*, "Microbial driven iron reduction affects arsenic transformation and transportation in soil-rice system," *Environmental Pollution*, vol. 260, p. 114010, 2020.
- [21] W. Zhai, T. Qin, L. Li, *et al.*, "Abundance and diversity of microbial arsenic biotransformation genes in the sludge of full-scale anaerobic digesters from a municipal wastewater treatment plant," *Environment International*, vol. 138, p. 105535, 2020.
- [22] S. Kumar, G. Stecher, M. Suleski, M. Sanderford, S. Sharma, and K. Tamura, "Mega12: Molecular evolutionary genetic analysis version 12 for adaptive and green computing," *Molecular Biology and Evolution*, vol. 41, no. 12, msae263, 2024.
- [23] J. A. Frank, C. I. Reich, S. Sharma, J. S. Weisbaum, B. A. Wilson, and G. J. Olsen, "Critical evaluation of two primers commonly used for amplification of bacterial 16S rrna genes," *Applied and environmental microbiology*, vol. 74, no. 8, pp. 2461–2470, 2008.
- [24] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of molecular biology*, vol. 215, no. 3, pp. 403–410, 1990.
- [25] C. L. Schoch, S. Ciufo, M. Domrachev, *et al.*, "Ncbi taxonomy: A comprehensive update on curation, resources and tools," *Database*, vol. 2020, baaa062, 2020.
- [26] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "Clustal w: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic acids research*, vol. 22, no. 22, pp. 4673–4680, 1994.
- [27] F. Sievers, A. Wilm, D. Dineen, *et al.*, "Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega," *Molecular systems biology*, vol. 7, no. 1, p. 539, 2011.
- [28] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, and T. L. Madden, "Primer-blast: A tool to design target-specific primers for polymerase chain reaction," *BMC bioinformatics*, vol. 13, pp. 1–11, 2012.
- [29] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch, "Expasy: The proteomics server for in-depth protein knowledge and analysis," *Nucleic acids research*, vol. 31, no. 13, pp. 3784–3788, 2003.
- [30] B. Bjellqvist, G. J. Hughes, C. Pasquali, *et al.*, "The focusing positions of polypeptides in immobilized ph gradients can be predicted from their amino acid sequences," *Electrophoresis*, vol. 14, no. 1, pp. 1023–1031, 1993.
- [31] B. Bjellqvist, B. Basse, E. Olsen, and J. E. Celis, "Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a ph scale where isoelectric points correlate with polypeptide compositions," *Electrophoresis*, vol. 15, no. 1, pp. 529–539, 1994.
- [32] E. Gasteiger, C. Hoogland, A. Gattiker, *et al.*, "Protein identification and analysis tools on the expasy server," *The proteomics protocols handbook*, pp. 571–607, 2005.
- [33] F. Teufel, J. J. Almagro Armenteros, A. R. Johansen, *et al.*, "Signalp 6.0 predicts all five types of signal peptides using protein language models," *Nature biotechnology*, vol. 40, no. 7, pp. 1023–1025, 2022.
- [34] J. Hallgren, K. D. Tsirigos, M. D. Pedersen, *et al.*, "Deeptmhmm predicts alpha and beta transmembrane proteins using deep neural networks," *biorxiv*, pp. 2022–04, 2022.
- [35] H. M. Berman, J. Westbrook, Z. Feng, *et al.*, "The protein data bank," *Nucleic acids research*, vol. 28, no. 1, pp. 235–242, 2000.
- [36] J. Jumper, R. Evans, A. Pritzel, *et al.*, "Highly accurate protein structure prediction with alphafold," *nature*, vol. 596, no. 7873, pp. 583–589, 2021.